



**FACULTY  
OF INFORMATION  
TECHNOLOGY  
CTU IN PRAGUE**

## ASSIGNMENT OF MASTER'S THESIS

**Title:** Recommendation based on product images  
**Student:** Bc. Kristýna Tauchmanová  
**Supervisor:** doc. Ing. Pavel Kordík, Ph.D.  
**Study Programme:** Informatics  
**Study Branch:** Knowledge Engineering  
**Department:** Department of Applied Mathematics  
**Validity:** Until the end of winter semester 2020/21

### Instructions

Survey recommender systems and focus on recommendation algorithms using product images (not exclusively). Design recommender systems for various scenarios such as homepage, related products, shopping cart or recommendation based on image provided. Evaluate performance of algorithms using offline data provided by Recombee. Discuss an improvement of recall and catalog coverage thanks to image embedding (if any).

### References

Will be provided by the supervisor.

Ing. Karel Klouda, Ph.D.  
Head of Department

doc. RNDr. Ing. Marcel Jiřina, Ph.D.  
Dean

Prague February 18, 2019





**FACULTY  
OF INFORMATION  
TECHNOLOGY  
CTU IN PRAGUE**

Master's thesis

## **Recommendation based on product images**

***Bc. Kristýna Tauchmanová***

Department of Applied Mathematics

Supervisor: doc. Ing. Pavel Kordík, Ph.D.

January 8, 2020



---

## Acknowledgements

Thank you...



---

## Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as school work under the provisions of Article 60(1) of the Act.

In Prague on January 8, 2020

.....

Czech Technical University in Prague

Faculty of Information Technology

© 2020 Kristýna Tauchmanová. All rights reserved.

*This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).*

### **Citation of this thesis**

Tauchmanová, Kristýna. *Recommendation based on product images*. Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2020.



---

# Abstrakt

**Klíčová slova** Doporučovací systémy, zpracování obrazu

---

# Abstract

**Keywords** Recommendation systems, image processing



---

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Literature review</b>	<b>3</b>
1.1 Content-based filtering . . . . .	3
1.2 Collaborative filtering . . . . .	5
1.3 Hybrid filtering . . . . .	6
<b>2 Analysis and design</b>	<b>9</b>
<b>3 Realisation</b>	<b>11</b>
<b>4 Experiments and discussion</b>	<b>13</b>
<b>Conclusion</b>	<b>15</b>
<b>Bibliography</b>	<b>17</b>



---

## List of Figures



---

# Introduction





---

# Literature review

When creating a recommending system which is going to be used on e-commerce sites, there are some scenarios that should be thought of. Different parts of an online store require different approaches so there are many recommendation systems specially tailored for a concrete online retailer or dataset.

According to cases and placement of the recommended products, we can meet with a group of items like “Customers who bought this also bought...”, “Items usually bought together”, “Bestsellers”, “Latest products”, “You might also like...”, “Similar items” and many others.

Even though the imagination in creating scenarios can be unlimited, approaches of designing a recommendation system can be divided into three basic categories - content-based filtering, collaborative filtering, and hybrid filtering.

## 1.1 Content-based filtering

Content-based filtering method focuses on items and their description. These recommendation systems are used when there is information about product (item is described by some features such as name, type, colour, manufacturer, etc.) and there is no information needed about the customer. Users (customers of a web store) are being tracked and according to their behaviour, the system decides to which items the user pays attention and so recommends items similar to users interest. For example, when a customer is looking at summer dress with a floral print, the system decides that this type of product is for the customer interesting and it is going to recommend other summer dresses with a floral print that look similar (have similar item description).

A typical application of this method is in “Similar items” or “You may also like...” sections of a web store.

One of recommendation system that uses image representation is introduced in [1]. The authors designed an algorithm and tested it on 600 images of clothes and shoes from the Amazon web store and JD web store. First of

all, they removed image background by modified Local Conditional Flooding algorithm and then extracted low-level image features - colour from HSV histogram, texture defined by Tamura features [2] and Gabor filters, and shape from geometrical descriptors. Then the features were normalized and their weight computed. Because the speed of response of the system was in this work important, the authors used an indexing method for quick searching for similar images in the dataset.

As well as in the previous article, researchers in [3] used for image representation low-level features. Histogram of Oriented Gradient (HOG), Shape Context [4] and Hu Moments were integrated together with text information (product text description was compressed by Long short-term memory method) and weighted to emphasize or suppress extracted features according to their importance. Authors collected 5000 products from online stores and evaluated their proposed algorithm. They considered a recommendation to be correct if all from top k recommended products were in the same category (e.g. glasses, shoes, watches etc.) as a query product. Cosine similarity was used for similarity calculation. They achieved 85 % of accuracy.

In article [5] a neural network classifier is used as a data-driven, visually-aware feature extractor. Batch-normalized Inception architecture was pre-trained on DeepFashion Attribute Prediction dataset ([6]) twice - for category and texture classification. Then it was applied on images from Fashion dataset, the output vectors from the pre-trained convolution neural networks were concatenated and used as image representations. For the top-k recommendations the authors used k-NN ranking algorithm.

The semantic gap between low-level and high-level features were tried to be solved by researchers in works [7] and [8]. They proposed image recommendation in vertical search based on ANOVA cosine similarity. They got image features (gray level co-occurrence matrix, Haralick features, Tamura features [2] and Gabor filters) and normalized them. For each term that appeared in the product text description, they computed ANOVA p-value by combining visual features and text-based search of the images and then used it as a weight of the term. Term dictionary of visual synonyms is then constructed according to term similarity. For user query, they generate expended queries using the dictionary and text-based search is performed. Images are recommended on cosine similarity score.

Another vertical image search is presented in [9]. The authors combine visual features with user relevance feedback. In the offline part of their proposed method, they extract five visual features for each image in a dataset - colour moments, colour correlogram, texture, local binary pattern and edge detection. In the online (recommendation) part, they retrieve user feedback from search history with clicked and unclicked images. The relevance of the images from the sets is decided on cosine similarity of the images features vectors and then they are ranked and recommended to the customer. They tested the method on crawled image data from myntra.com, user relevance feedback

was simulated with 100 participants. They evaluated their system manually and came to better accuracy of relevance score than with CBIR method.

The authors in article [10] proposed a novel representation of images called Visual Part-based Object Representation. The main principle of the method is to decompose a product image into a set of disjoint parts and let users say which parts of the product they care about. According to their preferences, the products similar to the currently displayed product are recommended with an emphasis on the selected area. In practice, when a customer is looking at a motorbike helmet, which is divided into top, shield and visor and chooses that she/he likes the shield part, the system will favour helmets with similar shields in its recommendations. The authors extracted low-level features from the separated parts (such as HSV colour histogram and Bag-of-Visual-Word histogram) and from the whole product image to represent each item. The proposed method was tested on images of 5 categories from Amazon web store and the authors came to a conclusion that it can achieve better performance than some text-based methods.

Researchers in [11] were focused on designing a user interface to browse fashion products comfortably. They presented a terminal for usage in retail stores where on the touch screen upper and lower parts of apparel (which are sold in the store) are shown. In the middle of the screen, there is currently selected outfit, around the clothes, there are similar pieces from the store recommended. The further from the selected outfit, the dissimilar products are recommended as an inspiration. Recommending works for upper and lower apparel separately. By clicking on another piece of clothes, the recommendations are recalculated and the screen is changed. For the product representation, they used image preprocessing (removing image background by use of Canny edge detection and dilatation) and then they extracted low-level image features such as correlated colour temperature, colour brightness and colour hue.

## 1.2 Collaborative filtering

Recommendation systems that use collaborative filtering method are based on human behaviour. They assume that people trust other people from their social environment and listen to their recommendations. In other words, the system will recommend products to customers according to a group of people with similar taste and interest (collaboration). For example, if the first person likes a backpack, hiking boots and hiking poles and the second person likes hiking boots, poles and a rain jacket, the system will recommend a rain jacket to the first person and a backpack to the second person.

This method analyzes the relationship between a customer and items and makes predictions about interests based on the similarity of users. Therefore collaborative filtering approach can recommend relevant items despite not

understanding the content of the sold products.

Common application of this method can be seen in “Customers who bought this also bought...”, “Items usually bought together” or “You might also like...” sections of a web store.

Article [12] managed to use collaborative filtering method together with image features. A rating prediction model was created where similar users helped to predict current user ratings. Then for all items, that meet a condition with a rating threshold, image features are extracted. Authors opted for Sparse filtering method to compute eigenvectors of a product picture. They calculated Euclidean distance between these items and already seen items of a user and the most similar products were recommended. They tested the proposed method on Amazon product dataset.

Another approach of combination collaborative filtering and image features can be seen in [13]. The basis of the predictor is matrix factorization which is combined with image features extracted by AlexNet convolutional neural network. They also took into consideration users’ behaviour over time and created a time-aware visual predictor. The model is then learned using Bayesian Personalized Ranking method.

The same approach chose researchers in their work [14].

In article [15] the researchers designed an image recommendation algorithm using feature-based collaborative filtering. Unlike in traditional collaborative filtering method, where a user-item matrix is used, in this work a user is represented by visual feature space. First of all, every image of a small dataset is segmented into several regions and visual features are computed for every region. For this purpose, they are described by low-level features such as color, texture and shape. Then the images that are purchased by the same user are clustered according to their features. A user is defined as a set of these clusters. Items that are recommended to a user depend on k-nearest neighbours that are calculated by inter-cluster distance.

### 1.3 Hybrid filtering

Combining content-based filtering and collaborative filtering methods results in hybrid filtering. This kind of technique can overcome problems of the previous two approaches and lead to better results. Merging can be done in various ways like creating two different models and combining the results at the end or by “spicing” one of the methods by adding some tactics of another.

With this definition, the application of hybrid filtering is almost limitless.

Researches in [16] proposed a hierarchical user interest mining method, which is based on user-contributed photos in her/his social media sites. They assume that photos shared by users on a web page have the same topic and use their textual description with the photos as an input to their model. They map user’s information to hierarchical topic space, ODP (open directory project),

which is a manually edited ontology directory. The item representation is mapped the same way. The user interest is calculated by the TF-IDF method and the relevance between a user and an item is measured by cosine similarity. This recommendation based on ODP is similar to collaborative filtering.

In [17] the recommendation system was divided into two stages - in the first stage, the system recommends products based on view-also-view model (classic collaborative filtering) and then the customer manually highlight regions of suggested items and the system re-ranks recommendations according to image features of the regions. The images are represented by concatenated low-level features such as colour and texture. Colour is extracted with the usage of HSV color space and separated into 30 colour bins. The texture is described by LBP (local binary patterns) and SIFT features. Image similarity is calculated with Kullback-Leibler divergence.

As well as the previous article, researchers in [18] focused on fashion. They designed a recommendation system capable of matching clothes and accessories together. The dataset they used is recorded from the Amazon web store. They downloaded images of products and recommendation shown there as relationships between them. For every image, they calculated its feature vector using a convolutional neural network - Caffe deep learning framework which is pre-trained on ImageNet. Then they learn a parameterized distance transform where for objects that are related the distance is smaller than to those that are not. Then the recommendations are made according to the category of a product and type of relation. They created a system that can recommend complementary products and one of the conclusion they came to was that James May is more fashionable than Richard Hammond.

Researchers in [19] also used convolutional neural networks to represent items and users. They designed a dual-net deep network, which takes triples as input - information about a customer and about two products. The CDL (comparative deep learning) architecture has three sub-networks - two are identical convolutional neural networks and extract features from input images, one is a full-connection neural network that captures user's information. CNNs extracting visual information of items are inspired by AlexNet and produce a 1024-dim vector representing input image. The output of the third sub-network is also a 1024-dim vector, but as an input, it takes user vector - firstly all possible tags from items are converted by word2vector method [20] into vectors, then these vectors are clustered and finally, users are described as bag-of-words applied on these clusters. Relative distances between the inputs are the objective of learning.

Article [21] focused on recommendations based on LSH (Locality Sensitive Hashing) algorithm. SIFT features are extracted from low-resolution images, then the descriptors are converted into eigenvectors and saved to the hashed table. The recommendation system takes an image as an input, returns target pictures (images from the same hash bucket) and with help from collaborative filtering recommends products to the user.

In [22] the authors proposed a hybrid recommendation system that uses aesthetic features of clothing. They trained an aesthetic neural network with AVA dataset [23], where they removed some pathways which have little connection to clothing and added low-level features of the images. The output of the second fully-connected layer was used as aesthetic features. During recommendation, both aesthetic features and CNN features (the second fully-connected layer of Caffe model trained on ImageNet dataset) contributed to the final prediction based on dynamic collaborative filtering. At the end, there was a tensor factorization model capable of better results than a model that used just CNN features.

Another recommendation system that combines CNN image features and collaborative filtering can be found in [24]. The researchers used the CNN model VGG19 pre-trained on ImageNet and performed classification of the products with the help of extracted image feature vectors. Then they used the same features to recommend products with respect to customer's preferences, her/his purchase history, rating of products and diversity of the product categories.

Also, authors of [25] used VGG19 neural network to extract high-level as well as low-level features. Then they found correlations between them with a usage of a convolutional neural network to express style features. All of these features were incorporated into a collaborative learning system based on Bayesian Personalized Ranking, which uses implicit feedback from users and predicts their interest.

# **Analysis and design**





# Realisation



## **Experiments and discussion**



---

## Conclusion



---

## Bibliography

- [1] Yu, L.; Han, F.; et al. A content-based goods image recommendation system. *Multimedia Tools and Applications*, volume 77, no. 4, Feb 2018: pp. 4155–4169, ISSN 1573-7721, doi:10.1007/s11042-017-4542-z. Available from: <https://doi.org/10.1007/s11042-017-4542-z>
- [2] Tamura, H.; Mori, S.; et al. Textural Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man, and Cybernetics*, volume 8, no. 6, June 1978: pp. 460–473, ISSN 2168-2909, doi: 10.1109/TSMC.1978.4309999.
- [3] Kawattikul, K. Product Recommendation using Image and Text Processing. In *2018 International Conference on Information Technology (IncIT)*, Oct 2018, pp. 1–4, doi:10.23919/INCIT.2018.8584860.
- [4] Belongie; Malik. Matching with shape contexts. In *2000 Proceedings Workshop on Content-based Access of Image and Video Libraries*, June 2000, pp. 20–26, doi:10.1109/IVL.2000.853834.
- [5] Tuinhof, H.; Pirker, C.; et al. Image-Based Fashion Product Recommendation with Deep Learning. *Lecture Notes in Computer Science*, 2019: pp. 472–481, ISSN 1611-3349, doi:10.1007/978-3-030-13709-0\_40. Available from: [http://dx.doi.org/10.1007/978-3-030-13709-0\\_40](http://dx.doi.org/10.1007/978-3-030-13709-0_40)
- [6] Liu, Z.; Luo, P.; et al. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, ISSN 1063-6919, pp. 1096–1104, doi:10.1109/CVPR.2016.124.
- [7] Sejal, D.; Ganeshsingh, T.; et al. Image Recommendation Based on ANOVA Cosine Similarity. *Procedia Computer Science*, volume 89, 2016: pp. 562–567, ISSN 1877-0509, doi:<https://doi.org/10.1016/j.procs.2016.06.091>. Available from: <http://www.sciencedirect.com/science/article/pii/S1877050916311565>

- [8] Sejal, D.; Ganeshsingh, T.; et al. ACSIR: ANOVA Cosine Similarity Image Recommendation in vertical search. *International Journal of Multimedia Information Retrieval*, volume 6, no. 2, Jun 2017: pp. 143–154, ISSN 2192-662X, doi:10.1007/s13735-017-0124-0. Available from: <https://doi.org/10.1007/s13735-017-0124-0>
- [9] Sejal, D.; Abhishek, D.; et al. IR\_URFS\_VF: image recommendation with user relevance feedback session and visual features in vertical image search. *International Journal of Multimedia Information Retrieval*, volume 5, no. 4, Nov 2016: pp. 255–264, ISSN 2192-662X, doi:10.1007/s13735-016-0111-x. Available from: <https://doi.org/10.1007/s13735-016-0111-x>
- [10] Chi, H.-Y.; Chen, C.-C.; et al. UbiShop: Commercial item recommendation using visual part-based object representation. *Multimedia Tools and Applications*, volume 75, no. 23, Dec 2016: pp. 16093–16115, ISSN 1573-7721, doi:10.1007/s11042-015-2916-7. Available from: <https://doi.org/10.1007/s11042-015-2916-7>
- [11] Piazza, A.; Zagel, C.; et al. Outfit Browser – An Image-data-driven User Interface for Self-service Systems in Fashion Stores. *Procedia Manufacturing*, volume 3, 2015: pp. 3521–3528, ISSN 2351-9789, doi: <https://doi.org/10.1016/j.promfg.2015.07.686>. Available from: <http://www.sciencedirect.com/science/article/pii/S2351978915006873>
- [12] Wang, Y.; Li, L. An Improved Personalized Recommendation Based on Purchasing Power and Browsed Images. In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, June 2018, pp. 321–328, doi:10.1109/HPCC/SmartCity/DSS.2018.00073.
- [13] Wu, Z.; Paul, A.; et al. Improved one-class collaborative filtering for online recommendation. In *2017 International Workshop on Complex Systems and Networks (IWCSN)*, Dec 2017, pp. 205–209, doi:10.1109/IWCSN.2017.8276528.
- [14] He, R.; McAuley, J. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. 2015, 1510.01784.
- [15] Kim, D.-H. Image Recommendation Algorithm Using Feature-Based Collaborative Filtering. *IEICE Transactions*, volume 92-D, 03 2009: pp. 413–421, doi:10.1587/transinf.E92.D.413.



- 
- [16] Feng, H.; Qian, X. Mining user-contributed photos for personalized product recommendation. *Neurocomputing*, volume 129, 2014: pp. 409–420, ISSN 0925-2312, doi:<https://doi.org/10.1016/j.neucom.2013.09.018>. Available from: <http://www.sciencedirect.com/science/article/pii/S0925231213009363>
- [17] Hsiao, J.; Li, L. On visual similarity based interactive product recommendation for online shopping. In *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, ISSN 1522-4880, pp. 3038–3041, doi:10.1109/ICIP.2014.7025614.
- [18] McAuley, J.; Targett, C.; et al. Image-Based Recommendations on Styles and Substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, New York, NY, USA: ACM, 2015, ISBN 978-1-4503-3621-5, pp. 43–52, doi:10.1145/2766462.2767755. Available from: <http://doi.acm.org/10.1145/2766462.2767755>
- [19] Lei, C.; Liu, D.; et al. Comparative Deep Learning of Hybrid Representations for Image Recommendations. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016, doi:10.1109/cvpr.2016.279. Available from: <http://dx.doi.org/10.1109/CVPR.2016.279>
- [20] Mikolov, T.; Chen, K.; et al. Efficient Estimation of Word Representations in Vector Space. 2013, 1301.3781.
- [21] Liu, D.; Huo, C.; et al. Research of commodity recommendation workflow based on LSH algorithm. *Multimedia Tools and Applications*, volume 78, 02 2018, doi:10.1007/s11042-018-5716-z.
- [22] Yu, W.; Zhang, H.; et al. Aesthetic-based Clothing Recommendation. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 2018, doi:10.1145/3178876.3186146. Available from: <http://dx.doi.org/10.1145/3178876.3186146>
- [23] Murray, N.; Marchesotti, L.; et al. AVA: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, ISSN 1063-6919, pp. 2408–2415, doi:10.1109/CVPR.2012.6247954.
- [24] Nelaturi, N.; Devi, G. L. Hybrid Recommender System Leveraging Stacked Convolutional Networks. *Journal of Engineering Science and Technology Review*, volume 11, 2018: pp. 89–96, ISSN 1791-2377, doi:10.25103/jestr.113.12. Available from: <http://dx.doi.org/10.25103/jestr.113.12>

## BIBLIOGRAPHY

---

- [25] He, M.; Zhang, S.; et al. Learning to Style-Aware Bayesian Personalized Ranking for Visual Recommendation. *IEEE Access*, volume 7, 2019: pp. 14198–14205, ISSN 2169-3536, doi:10.1109/ACCESS.2019.2892984.